厦门大学

XIAMEN UNIVERSITY

A STUDY ON INFLUENCE ANALYSIS OF TWO TYPES OF BIASED ESTIMATORS IN LINEAR REGRESSION MODELS

A Dissertation

By

ABOOBACKER JAHUFER

DOCTOR OF PHILOSOPHY (Ph.D)
IN
STATISTICS

May 2010

University Code: 10384 **UDC________**
Student Code: 15420070154132

A STUDY ON INFLUENCE ANALYSIS OF TWO TYPES OF BIASED ESTIMATORS IN LINEAR REGRESSION MODELS

A dissertation
By

ABOOBACKER JAHUFER
Student No: 15420070154132

Supervisor: Prof. JIAN-BAO CHEN

A dissertation
Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY (Ph.D)
IN
STATISTICS

School of Economics
Department of Planning and Statistics
Xiamen University
China

May 2010

学校编码：**10384**　　　　　　　　　　　　　　　　分类号______密级_____

学号：15420070154132　　　　　　　　　　　　　　　　　**UDC**_____

厦门大学

博　士　学　位　论　文

A Study on Influence Analysis of Two Types of Biased Estimators in Linear Regression Models

线性回归模型中两类有偏估计的影响分析研究

Aboobacker Jahufer

指导老师姓名：陈建宝 教授

专 业 名 称：统 计 学

论文提交日期：2010 年 3 月

论文答辩日期：2010 年　月

学位授予日期：2010 年　月

答辩委员会主席：____________

评　　阅　　人：____________

2010 年 05 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（Chinese Scholarship Council (CSC), China）课题（组）的研究成果，获得（CSC）课题（组）经费或实验室的资助，在（　　　　　　　）实验室完成。请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　月　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　年　月　日解密，解密后适用上述授权。

（　　）2.不保密，适用上述授权.

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权.）

声明人（签名）：

年　月　日

Dedicated to My Beloved
Parents and Family

ACKNOWLEDGEMENT

Charged with the blessing of Almighty GOD and the goodwill of the good, I have now come to complete the Ph.D study exampling my strength, patience and ability with the opportunity extended to me by the Chinese Scholarship Council (CSC) in collaboration with the Sri Lankan Government.

At this juncture of satisfaction in my endeavor towards carrying out this research study, I must first acknowledge my profound and sincere gratitude that I deeply owe to my supervisor Professor Jian-Bao Chen, Department of Planning and Statistics, School of Economics, Xiamen University for the full interest taken to persuade and motivate me with his rare talents exhibited through his valuable guidance, praise worthy comments, innovative suggestions, periodic advices, constant stimulations, useful discussion and fascinating supervision throughout the period of study, research and preparing compiling the research dissertation for my future professional peaks of popularity and potentiality.

I feel that it is my foremost duty to express my special thanks to the CSC, People's Republic of China for granting me the Ph.D scholarship - 2006 to undertake the Ph.D study in statistics, at the Department of Planning and Statistics, School of Economics, Xiamen University, Xiamen, P.R. of China.

I remain ever with deep gratitude the President of Xiamen university who was very kind enough to consider me for the admission to this Ph.D study at Xiamen university; Dean school of economics who was very enthusiastic in facilitating me to promote my studies; all professors and teachers who were cordially assisted me to accommodate my activities in this Ph.D studies with humility, all administrative staff who have extended their cooperative spites and goodwill to enable me to perform my mission of completing this study and the intimate classmates who were very affectionate and passionate to see I do my work peacefully in their presents.

Also, my appreciation should be extended to all the staff of Overseas Education Service Centre Office for their helps made available to me in magnificent manner to complete successfully the learning of one year Chinese language and to continue my Ph.D study and research in statistics in the Chinese environment.

It is my paramount duty and obligation as an academic to express my deep gratitude to the Vice-Chancellor, with Dean Faculty of Applied Sciences and the Head Department of Mathematical Sciences for granting me study leave and other necessary facilities as the mark of engorgement to enable me to complete the Ph.D degree programme in Statistics at Xiamen University, China.

I will be failing my duties if I miss extending my deep gratitude that I morally owe to the members of Council, Senate and Faculty Board of South Eastern University of Sri Lanka (SEUSL) for having kindly taken decision to enable me to enjoy the benefit of the awards the Chinese Scholarship for the betterment of my professional personality.

I also sincerely wish to express my gratitude to the Administrative staffs of SEUSL and my associate academic colleagues of SEUSL for assisting me in attending to my requirements for supplementing my request for the scholarship award.

I feel proud to express extra ordinary thanks with full of affection to Professor (Mrs.) P. Wijekoon, Dr. M.D. Mostafizur Rahman, Dr. A. Pushpanathan, Mr. S. Sivarajasingam, Mr. I.L. Fajurudeen and Mr. M. Razik to read this Ph.D degree programme.

I reserve boundless gratitude and indebtedness to my father, mother, father in law and all of my family members and relatives for inspiring and extending lot of moral assistants in the long process erecting my academic profile.

At last but not least I cannot find appropriate words to express enough and deep thanks to my lovely wife Firnaza and children Bassim Ahamed, Fathima Liba and Palah Leena for their support and encouragement shown considerably and numerous enthusiastic support to me to enable me to reach this significant peak of goal in my professional career.

The Author

摘 要

在过去的两世纪多，普通最小二乘法（OLS）得到了广泛的发展和应用，与之相关的研究领域包括物理学、医学、生物学、地质学、工程学、经济学、管理学和社会学等。自该方法使用以来，人们就普遍认为基于最小二乘法的统计推断对异常值具有敏感性，因此，诊断技术（如残差图等）被发展用来探测这些异常值和检测模型的假设是否正确等。

一般来讲，样本中的不同数据集会对拟合回归模型产生不同的影响。这也就是说，参数估计或者预测有可能更多的依靠样本中的小部分而不是大多数样本。这就需要指出这些对模型具有较强影响的样本点，并估算这些样本对模型的影响。如果这些具有较强影响的样本点本身存在问题，我们应该加以删除；另一方面，若这些点本身并没有什么问题，那么有必要对所建立的模型进行修正。

在线性回归建模过程中，当回归自变量间具有多重共线性时，尽管相应回归参数的 OLS 估计量仍然是最佳线性无偏估计(BLUE)，但是多重共线性却会使其产生很大的方差，从而导致其均方误差也很大，其后果在于相应的统计推断并不可靠。为了克服多重共线性性带来的后果，人们在大量文献中提出了约束最小二乘、混合估计和岭型有偏估计等。

作为解决多重共线性问题并且发展较为迅速的岭型有偏估计，具有很强的思想性和实用性，使用较为广泛，尤其是在计量经济学，工程技术和其他相关领域，岭型有偏估计分析已经为大量新理论、技术、思想和方法的发展打下了坚实的基础。

到目前为止，尽管不少岭型有偏估计量都可以用来拟合具有多重共线性数据的回归模型，但人们对这些有偏估计的影响分析方法并没有做出充分的研究。很多研究者都曾试图证明“观测值对岭型有偏估计的影响与其对相应 OLS 估计的影响有很大差异，岭型有偏估计可以冲淡异常数据对多重共线性的影响”，那么，结果到底如何呢？

本文试图在自变量存在多重共线性的条件下，将总体影响分析（Global Influence Analysis）、局部影响分析（Local Influence Analysis）、Cook 的最小扰动影响分析（Cook's Minor Perturbation Influence Analysis）以及 Box-Cox 变换影

响分析（Box-Cox Transformation Influence Analysis）等四种影响分析方法作进一步的发展，以便用于讨论两类有偏估计中的强影响观测值，这两类有偏估计分别是调整岭回归估计（Modified Ridge Regression Estimator，MRRE）和 Liu 估计。

在全局影响分析中，我们讨论了基于 MRRE 和 Liu 估计的残差、杠杆值、DFFITS 值以及 Cook 距离的两种形式，并导出了 DFFITS 和 Cook 距离的两种形式甄别影响点的近似删除公式。在局部影响分析中，我们通过对模型中的方差，解释变量进行扰动以及适当的选择岭参数估计，得到了两类有偏估计的广义 Cook 影响度量；在 Cook 最小扰动影响分析中，我们分别对误差方差是已知和未知的情况，讨论了两类有偏估计的影响分析。最后，在误差方差扰动的情况下，对 Box-Cox 变换模型的两类岭估计进行了影响分析研究。

为了具体说明 MRRE 和 Liu 估计的总体影响分析方法、局部影响分析方法、Cook 最小扰动影响分析方法以及 Box-Cox 转化影响分析方法，本文利用两组宏观经济数据进行了实例分析。

关键词: 调整岭回归估计; Liu 估计; 影响分析; 多重共线性; 强影响观测值

ABSTRACT

In the past two centuries, the ordinary least squares (OLS) method has in use. It is being developed and applied in many research fields, the related research fields are Physics, Medicine, Biology, Geology, Engineering, Economics, Management Science, Sociology and so on. From the beginning, it was generally believed that the statistical inference based on least squares is sensitive to unusual data. Therefore, some diagnostic techniques (for example, residual plot) were developed to identify anomalous observations and check violations of model assumptions.

In general, different subsets of the data may produce different influence on the fitted regression model. That is, parameter estimates or predictions may depend more on the influential points than on the majority of the data. It would be liked to locate these influential points and assess their impact on statistical influence of the model. If these influential points are bad values then they should be eliminated. On the other hand, there may be nothing wrong with these points, but if they control key model properties, as they would like for them to, they could affect the use of the model. Therefore, it is necessary to modify the model.

In the linear regression model building process, if there exists multicollinearity among the regressors, the high variance of OLS estimator can make the corresponding mean squared error (MSE) very large although, it is still a best linear unbiased estimator (BLUE). This may lead to the statistical inference unreliable. To overcome the multicollinearity problems, the restricted least squares (RLS), mixed estimators and ridge type biased estimators methods are given prominence in literature.

The ridge type biased estimator models have become an ideal, popular and useful technique in the path of its fast growth in the world of knowledge to solve multicollinearity problems. Also the recent literature, particularly touching upon the fields of econometrics, engineering and other statistical areas the ridge type biased estimators analysis, have come to lay concrete foundation to produced a number of new theories, techniques, ideas and methods.

Although many ridge type biased estimators have been proposed to fit the regression model for multicollinearity data, so far, very few related methods on influential analysis for ridge type biased estimators have been developed. Many authors have attempted to prove that “the influence of the observations on ridge type biased estimators is different from that of the corresponding OLS estimator. Whether this statement is true or false?

In this research study, it is aimed to further develop influence analysis methods including global, local, Cook's minor perturbation and Box-Cox transformation to detect influential observations in ridge type biased estimators namely modified ridge regression estimator (MRRE) and Liu estimator when these two estimators are used to mitigate the effects of multicollinearity is existing among the regressors.

Under the global influential analysis residual, leverage, DFFITS and two versions of Cook's distance are studied for MRRE and Liu estimator. In addition to these influential analyses for both ridge type estimators, approximate deletion formulas are proposed for detection of influential points using DFFITS and two versions of Cook's distance. Moreover, under the local influential analysis variance perturbation, perturbation of individual explanatory variable and an assessment of biasing parameter are analyzed for both biased estimators. Under the Cook's minor perturbation influential analysis variance perturbation when variance of error is known and unknown situations and perturbation of individual explanatory variable are studied for both ridge type estimators. Finally under the Box-Cox transformation variance perturbation is studied for both biased estimators.

To illustrate the methodologies developed under the global, local, Cook's minor perturbation and Box-Cox transformation influential analysis for MRRE and Liu estimator, two real macroeconomic data sets are used.

Key Words: Modified Ridge Regression Estimator; Liu Estimator; Influence Analysis; Multicollinearity; Influential Observation.

CONTENTS